W-043E
C.2

THE METHODOLOGY OF THE CANADIAN AIR SCHEDULED INTERNATIONAL PASSENGER ORIGIN AND DESTINATION ESTIMATION SYSTEM

Greg Hunter and Lisa DiPiétro

Statistics Canada

Prepared for presentation at
The American Statistical Association,
Cincinnati, Ohio
August 1982

W-043E
C.2.

THE METHODOLOGY OF THE CANADIAN AIR SCHEDULED INTERNATIONAL PASSENGER ORIGIN AND DESTINATION ESTIMATION SYSTEM

Greg Hunter and Lisa DiPiétro

Statistics Canada

INTRODUCTION

The Air Scheduled International Passenger Origin and Destination (ASIPOD) estimation system uses the data from two air traffic surveys to produce estimates of the number of passengers on scheduled international flights between Canadian and foreign markets for various origin-destination combinations. The first of these two surveys, the revenue passenger origin and destination survey, provides a sample of origin-destination data on international journeys with Canadian carriers on at least one leg of the itinerary. A coverage problem exists, since no data are available on those international journeys with foreign carriers on all legs of the itinerary. The second survey, the airport activity survey, counts all passengers entering or leaving Canada on all Canadian or foreign scheduled carriers without consideration of the passenger's origin or destination.

This paper first outlines the requirements users have for international passenger origin and destination estimates. Then, the relevant aspects of the two air traffic surveys and the non-coverage[1] problem are presented. And, finally, the paper describes how the ASIPOD estimation system will produce estimates of the number of passengers and the associated coefficients of variation for various international origin-destination pairs for the portions of the international market both covered and not covered by the first survey.

BACKGROUND

The redesign of the ASIPOD estimation system is one of several projects in a revision of the federal aviation statistics program. In 1979, Statistics Canada embarked on the revision program by inviting Transport Canada and the Canadian Transport Commission (i.e. the two "user departments") to form an interdepartmental revision team, the Aviation Statistics Requirements Committee (ASRC). The ASRC acts as a senior review group for the ASIPOD project team, which consists of a project manager, representatives of the user departments and specialists in mathematical statistics, aviation economics and data processing. Thus far, the project team has specified the user requirements, undertaken feasibility work, written specifications for the methods and designed the computing system.

Ce document est aussi disponible en français[2].

USER REQUIREMENTS[1],[2]

The ASIPOD project team identified user requirements as described in the following two sections.

(1) The need for estimates of international passengers by origin and destination.

Before an international scheduled commercial air service[3] can be operated into and out of Canada, some form of formal agreement must exist between the Government of Canada and the government of the second country. The formal agreement between countries may take the form of an interim diplomatic exchange of notes or of a complete negotiated Air Transport Agreement.

International bilateral air negotiations involve officials of the Canadian government from External Affairs, the Canadian Transport Commission, Transport Canada and the Ministry of Industry, Trade and Commerce. Such negotiations may last several months or several years.

The routes for scheduled air services are normally the major item for negotiation, but there are many others. Some of the items, or articles written into air transport agreements, may be:

- rights to fly across, or to make stops for non-traffic purposes in, a given territory
- designation of the airline to operate each route
- compliance with laws and regulations of each country, dealing with such issues as entry, clearance, immigration, passports, customs and quarantine
- airworthiness, certificates of competency, and licences
- exchange of statistical information
- tariffs
- transfer of funds
- exemption from taxation of income

In order to negotiate for these items, and particularly for an exchange of routes, the negotiating officials must know where air traffic markets are and whether they are growing. Analyses of the costs and benefits to Canada and to foreign countries of various international routes for Canadian and foreign air carriers must be available to the negotiating officials. To ensure that Canada can negotiate a fair market share, the provision of international

passenger origin and destination estimates is a necessity. A crude and indirect indication of the value of such data to the Canadian economy is that the revenue generated from all international air routes to and from Canada in 1980 was about 2.3 billion dollars (Canadian).

(2) The objectives of the revision of the estimation system.

The four main objectives, the fourth of which will be discussed in detail in this paper, are as follows:

(i) to eliminate problems which have been identified in the existing system.

The existing system does not impute for illegible carrier codes on flight coupons, so that "unknown carriers" becomes the third largest carrier in tabulations. Also, there is no check on the coding which indicates whether a carrier is flying to and from airports at which it actually has landing rights. Edits and imputations to handle illegible or incorrect data on flight coupons are required over and above those which would be required for purely domestic tabulations.

Other nonsampling errors have been identified, but can not be easily corrected by an estimation system. Examples of such errors are misinterpretation by participating carriers of instructions for selecting flight coupons; systematic errors in serial numbers, used for sample selection, on ticket stock; errors in the carriers' processing systems, etc. The correction of these nonsampling errors for which it is not easy to correct is not an objective of the revision.

(ii) to develop a simple computing system which is easy to use, and produces summary diagnostic information.

A large volume of records[4] must be processed with minimal manual intervention. Since large volumes of passengers are dispersed across a large number of origin-destination pairs, the diagnostics at each stage of the system must summarize the processing, and still be able to point out potential problems.

(iii) to tabulate the international passenger estimates, regularly and on an ad hoc basis, in ways which will simplify the analyses undertaken by users.

(iv) to produce quantifiably reliable and accurate estimates of the number of air scheduled international passengers by origin and destination.

It has been thought that the accuracy of these statistics has been variable in the past. However, the reliability (i.e. sampling variability) of these statistics has been, in fact, unknown to date. Estimating the reliability of these data will improve the knowledge of the strength of the inferences that can be made from analyses of these data. Inferences made without knowledge of the reliability of the data could actually be quite misleading.

FEASIBILITY PHASE[3]

An underlying assumption of the feasibility work is that the same basic methodology, as implemented in the existing system, is to be used. As a result, most of the feasibility work involved identifying desirable improvements, ranking their desirability and determining how much could be done within cost and time constraints. This "same basic methodology" provided direction with respect to the calculation of estimates of the number of international passengers, but not to the calculation of estimates of reliability of the passenger estimates.

The target population of the international estimation system is the set of all tickets with an international (i.e. between a foreign country and either Canada or the United States) journey. The revenue passenger origin and destination survey collects tickets issued for international journeys, but only major Canadian carriers participate in the survey[5]. Each participating carrier selects a flight coupon on a ticket with a serial number ending in '0', if that carrier is the first participating carrier to fly on a leg of that ticket. Hence, the survey reports a 10 percent sample of unique flight coupons on which there is at least one participating Canadian or American carrier.

Some information concerning the markets of foreign (non-U.S.) carriers is obtained from the revenue passenger origin and destination survey. For example, if a passenger travels from Ottawa to Montréal with Air Canada, then connects with Air France for Paris, the revenue passenger origin and destination survey will capture the trip because a Canadian carrier participated somewhere in the journey. The Canadian carrier would report the complete carrier and routing detail, including the Air France segment.

The revenue passenger origin and destination survey, however, does not cover coupons with foreign carriers on <u>all</u> legs of an itinerary. An example of such an itinerary would be that of a passenger flying on Air France from Paris to Montréal and then back to Paris on Air France. If this itinerary were the passenger's total journey, this journey would not be reported to the revenue passenger origin and destination survey. However, the itineraries of such passengers <u>are</u> in the target population of international journeys.

This incomplete coverage of the target population is the non-coverage problem for the ASIPOD estimation system.

The existing system takes the revenue passenger origin and destination survey data and the airport activity survey data and applies a method called "the assignment technique" in order to produce total market estimates.

The airport activity survey counts passengers, on a census basis by flight, entering and leaving each Canadian airport. The survey covers all Canadian, American and foreign scheduled carriers, but it does not consider the passengers' initial origin or final destination.

Hence, the airport activity survey provides a count of the total volume of passengers for all carriers in the target population. The assignment technique is a method of estimating the non-coverage volume of passengers and assigning it to origin-destination pairs. However, in order to explain the assignment technique, a somewhat more thorough description of the two air traffic surveys is required.

RELEVANT ASPECTS OF THE TWO SURVEYS

The authorizing agency for the two survey programs is the Air Transport Committee of the Canadian Transport Commission, in co-operation with Transport Canada. The data are collected from the air carriers on behalf of the Air Transport Committee by the Aviation Statistics Centre (ASC) of Statistics Canada. Under the authority of the Air Carrier Regulations of the Aeronautics Act, reporting by the carriers on the ASC statements (ie. questionnaires) is compulsory.

The Revenue Passenger Origin and Destination (O & D) statistics are reported to the ASC via Statement 55. The reported data items, among others, include:

- ticket origin and ticket destination
- points of intra- and interlining (i.e. routing)
- carrier on each flight coupon stage

The revenue passenger origin and destination data are submitted monthly by major Canadian unit toll air carriers conducting scheduled passenger services. Since January 1, 1982 the seven Canadian carriers contributing information to this survey have been Air Canada, CP Air, Eastern Provincial Airways, Nordair, Pacific Western Airlines,Air Ontario and Quebecair. The American data are collected by the Civil Aeronautics Board from all certificated United States' air carriers, except from helicopter operators and intra-Alaska carriers. The data for the three months of each quarter are combined, and duplicates are eliminated; so that a file of complete itineraries, Ticket Origin and Destination (TOD) records, is obtained for the quarter.

However, passenger origin and destination statistics are compiled using the Directional Origin and Destination (DOD) concept. The DOD concept can be defined as "points of initial departure and ultimate destination named in the sequence which indicates the direction of travel". DOD's are pieces of itineraries which are broken up such that each component piece defines a reasonably consistent direction. To create DOD's, "open-jaw" and return itineraries, such as "symmetrical" and "circle" itineraries, must be broken into pieces which are essentially one-way trips. To obtain the DOD's, the TOD's are passed through the breakpoint routines. This breakpoint process is automated within the Passenger Origin and Destination System, and involves the calculation of various point-to-point distances within the itinerary and the comparison of these distances to the total itinerary length. As a general rule, itineraries are broken at the farthest point from the origin. Each DOD formed is recycled through breakpoint routines until no further breakpoints can be assigned.

The airport activity data are filed on statement 32. The relevant items included for each flight are:

- the reporting carrier
- the reporting airport
- the point of origin and final scheduled destination of the flight
- the last station arrived from, for arrivals; or next station departed to, for departures
- the number of deplaned or enplaned revenue passengers

The airport activity data are submitted monthly by Canada's transcontinental (Air Canada and CP Air) and regional air carriers (Eastern Provincial, Quebecair, Nordair and Pacific Western Airlines), by Norcanair and by all foreign carriers (including American carriers) operating scheduled international flights into and out of Canada. Since January 1, 1982, there have been 10 American and 21 other foreign carriers filing reports for each Canadian airport they served. Each new foreign carrier, granted a licence to serve Canada on a scheduled basis, is automatically included as a participant in the airport activity data collection system.

From the airport activity data, the census traffic flow data are obtained. "Traffic flow" can be defined as a count, over a certain period of time, of the number of persons who are flying on a specific carrier between a Canadian reporting airport and an adjacent point. The adjacent point is called the next stop or the last stop. For the purposes of the assignment technique only the traffic data for foreign (non-U.S.) carriers are input into the system. The data elements extracted from this survey, and used to determine the O & D international markets, are the number of revenue passengers enplaned and deplaned in Canada, the Canadian gateway carrier, and the Canadian gateway. In this survey the concept of "Canadian gateway" is defined to be that reporting airport at which a foreign (non-U.S.) carrier enters or leaves Canada.

However, in the revenue passenger O & D survey, the Canadian gateway for Canadian and U.S. carriers is the first/last Canadian point in the itinerary for a flight entering/leaving Canada. For foreign carriers the Canadian gateway refers to the point inside Canada where the passenger enters or leaves the foreign carrier. Consider the following fictitious example:

Assume that Air France flies Toronto - Montréal - Paris. Some passengers enplaned in Toronto, and some enplaned in Montréal.

Assume also that the matching single crossing DOD's are:

Winnipeg - Air Canada - Toronto - Air France - Montréal - Air France - Paris.

Toronto - Air France - Montréal - Air France - Paris - British Airways - London

The Canadian gateway in the above example would be Toronto because Toronto is the point at which the passengers enter the foreign carrier.

SOLUTION OF THE NON-COVERAGE PROBLEM

The non-coverage problem will be examined in four stages:

(1) Magnitude of the Problem

As in other surveys, non-coverage is a problem for the ASIPOD methodology, since no sample data are available on the origin-destination patterns of the non-coverage portion of the target population.

The following table gives an indication of the volume of passengers travelling between Canada and nine world areas. (These data are 1979 annual estimates. The world areas are not identified because these data are confidential.)

Table 1 - Estimated Number of International Passengers - 1979

Between Canada and ...	Revenue Passenger Origin and Destination	Non-coverage (percentage of Total in brackets)		Total
World Area #1	116,050	495	(0.4)	116,545
World Area #2	325,200	2,020	(0.6)	327,220
World Area #3	99,010	12,628	(11.3)	111,638
World Area #4	67,200	14,558	(17.8)	81,758
World Area #5	575,010	126,076	(18.2)	703,086
World Area #6	205,180	101,464	(33.1)	306,644
World Area #7	1,221,040	908,876	(42.7)	2,129,916
World Area #8	54,410	56,807	(51.1)	111,217
World Area #9	45,330	57,267	(55.8)	102,597
Total World	2,708,430	1,282,191	(32.1)	3,990,621

From the percentage non-coverage figures, it is evident that the non-coverage problem is a major concern.

The same table as above, but between Eastern Canada and the same nine world areas, would tend to have a higher percentage non-coverage for each world area. Hence, a lower level of geographic aggregation in origin-destination pairs generally implies a higher percentage non-coverage. For example, the non-coverage for Eastern Canada to World Area #7 is 55 percent, compared to the 42.7 percent tabulated for all of Canada to World Area #7 (as above). To clarify the idea that a higher level of geographic aggregation in origin-destination pairs generally implies a lower percentage non-coverage, consider the fact that non-coverage exists only for foreign traffic terminating at Canadian gateways. There is complete coverage of all traffic for which the Canadian end of the origin-destination pair is not a Canadian gateway. Hence, as the level of geographic aggregation becomes higher, more and more interlining traffic is included, and the percentage non-coverage becomes lower.

(2) The Assignment Technique[4]

The assignment technique estimates the non-coverage volume of passengers and then allocates this volume to origin-destination pairs.

From the airport activity survey, b benchmark counts of passengers entering and leaving Canada at Canadian gateway airports are tabulated by carrier. The value of b, the number of assignment groups, can be determined as follows:

$$b = 2 \times \sum_{i=1}^{g} n_i$$

where the '2' accounts for the fact that there is one count each for passengers entering and leaving Canada,

g is the number of Canadian gateway airports, and

n_i is the number of foreign (non-U.S.) carriers with landing rights at the i^{th} Canadian gateway airport.

From the revenue passenger origin and destination survey, a corresponding number of inbound and outbound passengers on international DOD's can be tabulated by crossing carrier and Canadian gateway airport. Hence, there are also b such counts from the sample survey data.

In the first stage of the assignment technique the non-coverage volume, A_i, of passengers in assignment grouping i can be estimated as follows:

$$A_i = C_i - (1/f) \times D_i \quad (i = 1,\ldots b).$$

where C_i is the airport activity census count in assignment grouping i.

f is the revenue passenger O & D survey sampling fraction (i.e. 1/10).

D_i is the sample number of international passengers in assignment grouping i.

A_i is, then, an estimate of the number of passengers, carried on the foreign carrier in the i^{th} of b assignment groupings, for which there is no origin-destination information.

The next stage allocates the non-coverage volume to origin-destination pairs. Such pairs in the non-coverage portion are called non-interlining DOD's, since they are DOD's, flown by foreign carriers, which do not interline with a participating Canadian carrier. The assignment technique imputes non-interlining DOD's in the i^{th} assignment grouping as follows:

(i) All of the DOD's contributing passengers to D_i are identified. (These would be DOD's which match on Canadian gateway, foreign carrier and direction.)

(ii) The domestic portion of such DOD's (i.e. the portion from the Canadian point to the Canadian gateway city) is eliminated. (The domestic portion of such DOD's would be on a Canadian carrier, and, therefore, would be picked up in the revenue passenger O & D survey. The resultant "truncated DOD's" are, then, non-interlining.)

(iii) The non-coverage volume, i.e. A_i, is assigned to the resultant "sample DOD's" in proportion to their original contribution to D_i.

New DOD records consisting of "assigned passengers" are produced.

The assignment technique assumes that the truncated DOD's are representative of the non-interlining traffic. As a result, some original sample DOD's are receiving more weight than they would in the revenue passenger O & D survey alone. Hence, the estimator, $\hat{d}_j^T$, for the total market number of passengers for the j^{th} origin-destination pair can be derived by adjusting the sampling fraction as follows:

$$\hat{d}_j^T = (1/f_j) \times d_j$$

$$\text{where } f_j = \frac{d_j}{(d_j/f) + a_j} \qquad (1)$$

and where f_j is the adjusted sampling fraction associated with the j^{th} origin-destination pair, from the revenue passenger O & D sample survey.

d_j is the sample number of international passengers in the j^{th} origin-destination pair, from the revenue passenger O & D sample survey.

a_j is the number of passengers assigned to the j^{th} origin-destination pair.

Note that a_j, according to point (iii) above,

is $$a_j = \frac{d_j}{D_i} \times A_i \qquad j \epsilon i$$

where $$\sum_{j \epsilon i} d_j = D_i$$

and $$\sum_{j \epsilon i} a_j = A_i$$

(3) An Example of the Assignment Technique

A simple example will illustrate how the assignment technique works. Assume that the airport activity data from British Airways indicated that 120 passengers enplaned at Montréal and went to London. Therefore, C_i = 120. Assume that the only two DOD's for this assignment grouping from the revenue passenger origin and destination survey are:

Sample Number of Passengers	Estimate of the Number of Passengers	DOD's
1	10	YWG-AC-YMX-BA-LON-LO-WAW
2	20	YYZ-CP-YMX-BA-LON-LH-HAM

where the codes are to be interpreted as:

Code	Denotes
YWG	Winnipeg
YMX	Montréal (Mirabel)
LON	London
WAW	Warsaw
YYZ	Toronto
HAM	Hamburg
AC	Air Canada
BA	British Airways
LO	LOT
CP	CP Air
LH	Lufthansa

Therefore, $D_i = 3$, and $A_i = 120 - (1/.1) \times 3 = 90$.

The truncated DOD's, the proportion of their contribution to D_i, the resultant number of assigned passengers and total market estimates are then:

Truncated DOD's	Proportions	Assigned Passengers(a_j)
YMX - BA - LON - LO - WAW	1/3	30
YMX - BA - LON - LH - HAM	2/3	60

(4) Shortcomings of the Assignment Technique

The assignment technique has some recognized shortcomings.

The basic assumption of the assignment technique is that the truncated DOD's are representative of the non-interlining traffic. Consider the hypothetical example above in order to determine whether this assumption is intuitively reasonable. The assignment technique presumes that passengers flying

on a particular foreign air carrier and originating in Montréal would have the same ultimate destination as passengers originating in Winnipeg or Toronto who fly through Montréal. Hence, the travel patterns of ethnic communities, for example the Polish and German communities in Toronto and Winnipeg respectively, might be used to impute for travel patterns of the more predominantly French communities in Montréal. And, in fact, the assumption that interlining and non-interlining travel patterns are the same was proven empirically to be suspect in a pilot test[5] conducted by the Canadian Transport Commission in 1977. Therefore, there is not only some intuitive but also some empirical evidence against the basic assumption of the assignment technique.

The accuracy of the estimates of the number of passengers by origin-destination pair is jeopardized by any violations of the assumption that truncated DOD's are representative of non-interlining traffic. Large volumes of passengers are allocated to origin-destination pairs, as was seen in Table 1 above, based on a small "effective" sample size. For example, the "effective" sampling fraction of the Canada - World Area #7 market is, not 10% as in the domestic survey, but 5.7% (ie. (100% - 42.7%) x 10%), because of the non-coverage of non-interlining traffic. Hence, a smaller than 10% sample of DOD's is used to allocate a large volume of passengers. Violations of the aforementioned assumption, then, would cause a potentially large bias in the estimates.

Since airport activity census counts are used as benchmark figures for traffic volumes, their accuracy is very important. Although no evaluations have been undertaken to investigate the magnitude of bias from nonsampling errors in the airport activity census counts, aviation statistics economists feel that this is a survey in which such errors would be small. There are currently, however, ongoing discussions with the major Canadian air carriers on how the reporting requirements of government agencies can be minimized. As a result of these discussions, the airport activity survey could become a sample survey. If a sample is to be designed, the accuracy of the activity counts of gateway airport passengers would be an important design consideration.

The assignment technique also assumes that <u>all</u> of the non-coverage is accounted for by the non-interlining traffic. An interesting way of validating this assumption would be to compare, for the same reference period, the airport activity census count for a <u>Canadian</u> air carrier to the analogous estimate from the origin-destination sample survey. This analogous estimate would be the sum of all passengers on DOD's with the same Canadian gateway airport and crossing carrier. It would be necessary to be able to determine whether differences in the estimates were ascribable to differences in the concepts of the two surveys, and if so, and whether these differences have been taken into account in the ASIPOD estimation system. If such differences have not been accounted for, then it could be that there is a problem with the assumption that all of the non-coverage is accounted for by non-interlining traffic.

Many of these shortcomings should be investigated. However, there are major problems in the existing system, and no alternative solutions, which are superior to the same basic methodology of the assignment technique and which can be implemented within time and cost constraints, have been found. Furthermore, the improvements in the estimates of the number of passengers by air traffic market in the new ASIPOD system will be substantial.

The use of the assignment technique to estimate for international markets is an innovative solution to a large problem. It does not completely solve the non-coverage problem, but it is a major step in the right direction, as will be the production of estimates of the variance of the estimates. These variance estimates should take account of the assignment technique and its assumptions, and, at the same time, give a meaningful measure of the reliability of the DOD estimates.

ESTIMATION OF VARIANCE[6]

The estimator of the variance of the international origin-destination estimates is a simple extension of the variance estimator for the revenue passenger origin and destination survey.

(1) Variance of the Estimate of Interlining Traffic

The development of the estimator of the variance of the revenue passenger origin-destination estimates is dependent upon the way in which tickets contribute passengers to origin-destination pairs (i.e. the domains of interest). Recall that each ticket is selected with probability 0.1, and that each ticket may be broken up into several segments or DOD's. Each ticket may contribute 0,1,2, etc. passengers to a given domain of interest. For example, the itinerary

YWG - AC - LON - BA - YYZ

would be broken, via the breakpoint routines, into the two DOD's

YWG - AC - LON
and
LON - BA - YYZ.

Consider the inbound plus outbound estimates which are total passenger figures, independent of direction. For such estimates this ticket would add passengers to, among others, the following domains of interest:

Domains of Interest	Passenger Count
Winnipeg - London	1
Toronto - London	1
Canada - London	2
Canada - Europe	2
Eastern Canada - Europe	1
Western Canada - Europe	1

Note that the number of passengers per ticket depends on the geographic level of aggregation, and, therefore, on the particular origin-destination estimate (i.e. domain of interest).

The estimate, $\hat{d}$, of the number of passengers from the revenue passenger origin and destination survey can be developed for a particular domain of interest as follows:

$$\hat{d} = \sum_{i=1}^{n} \frac{x_i}{f}$$

where x_i is the number of DOD's belonging to the domain of interest on the i^{th} ticket.

n is the number of sample tickets selected, and

N is the population number of tickets in the revenue passenger origin and destination survey.

f is the sampling fraction, i.e. $n/N = .1$

The revenue passenger origin and destination survey sample is effectively a 10% simple random sample because

(i) the selection of coupons with serial numbers ending in the digit '0' produces a systematic sample, and

(ii) there is no cycle associated with the distribution of tickets which would cause a relationship between the survey estimates and the last digit of the serial number.

The estimate of the variance can be written, then, as

$$\hat{var}(\hat{d}) = N^2 \frac{1}{n} (1 - f) v_s$$

$$\text{where } v_s = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\text{and } \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad ,$$

and the coefficient of variation, as

$$\hat{cv}(\hat{d}) = \sqrt{\hat{var}(\hat{d})}/\hat{d} \quad .$$

Note that var(d) can also be written as

$$var(d) = \frac{(1-f)}{f^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (2)$$

where n is assumed to be sufficiently large for n/(n-1) to be approximately equal to 1.

(2) Variance of Total Market Estimates

The method for calculating total market coefficients of variation has to recognize that the assigned data are a function of the sample data. In other words different samples will produce different assigned data and, thereby, different values of the total market estimate. The re-use of certain portions of the sample has an effect on the sampling distribution of the estimates.

The method which will be used adjusts the sampling fraction from 10% to be the percentage for which sampled records for a particular domain of interest are actually accounting. Since the use of the assignment technique is a given, it has to be assumed that truncated DOD's from the revenue passenger origin and destination survey are representative of the non-interlining traffic. The measure of reliability will be a measure of the precision of the DOD estimates, only to the extent to which this assumption is valid. As a result, it is reasonable to adjust the weights of the sample DOD's to take into account the non-interlining DOD's. The sampling fraction, then, for the estimation for the j^{th} domain of interest would be fj, as developed in equation (1) above. fj would replace f in the formula for var(d) in equation (2) above in order to yield the formula for the variance of the total market estimates:

$$\hat{var}(\hat{d}_j^T) = \frac{(1-f_j)}{f_j^2} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

And, the coefficient of variation for the total market estimate would be

$$\hat{cv}(\hat{d}_j^T) = \sqrt{\hat{var}(\hat{d}_j^T)} \; / \; \hat{d}_j^T \quad .$$

Note that $f_j < .10$ for $a_j > 0$. This means that less than a 10% sample is achieved when it is necessary to include assigned data in a total market estimate. Hence, by using f_j instead of .10 in the expression for the variance of a DOD estimate, the coefficient of variation is adjusted to relate to the total market estimate.

This method gives credit to the use of sample data in the assignment technique; but it is dependent, as is the determination of the estimates themselves, upon the assumption that truncated DOD's are representative of non-interlining traffic.

FUTURE CONSIDERATIONS

Earlier, the need for data on air traffic markets in international bilateral air negotiations was explained. The exchange of statistical information is one of the negotiable articles in air transport agreements. Currently, agreements for the exchange of statistical information exist with several countries. The concepts and quality of the data from some of these countries indicate that these data could be used in the ASIPOD estimation system. Such data would provide sample information on the non-coverage portion of the international universe of tickets. Feasibility work is currently underway to determine whether the number of countries for which these data can be used would improve the accuracy of a sufficient number of estimates to justify the expansion of the ASIPOD system to use "exchange data".

SUMMARY

The ASIPOD estimation system, then, uses the data from two air traffic surveys to produce origin-destination estimates of international passengers. The assignment technique is the solution to the problem caused by the non-coverage of non-interlining traffic. The assumptions of the technique are sufficiently questionable to warrant an evaluation of the bias of the estimates. However, major improvements will be made in the new system which will decrease the bias in the estimates. Also, estimates of reliability will be produced. And as a result, knowledge of the strength of the inferences made with respect to air traffic markets from these estimates will be improved in international bilateral air negotiations.

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues in the Transportation and Communications Division and the Business Survey Methods Division of Statistics Canada for their many helpful comments during the development of this paper.

FOOTNOTES

1 The coverage problem seems to be "non-coverage" as opposed to "under-coverage", since it is not even possible to include a large portion of the universe in the frame.

2 Pour recevoir une copie de la version française, veuillez écrire à:

Greg Hunter,
Division des méthodes d'enquêtes-entreprises,
Édifice Jean Talon, 5ième étage, Section C8,
Tunney's Pasture,
Ottawa, Ontario, Canada,
K1A 0T6

3 An international scheduled commercial air service is defined to be an operation which is between points in Canada and points in any other country, and which provides public transportation of persons, goods or mail by aircraft in accordance with a schedule and at a toll or charge per unit of traffic. Such a service is referred to as a "unit toll" service.

4 There are some 625,000 origin and destination records annually, and some 820,000 airport activity records annually.

5 An exchange program on passenger O & D data is maintained between Canada and the United States, whereby the United States gives Canada those records detailing the complete itineraries of the tickets collected in their survey on which

(i) a U.S. and Canadian point is shown in the routing, or
(ii) a U.S. carrier is recorded as having flown to or from a Canadian point, or
(iii) a Canadian carrier is recorded as having flown to or from a U.S. point.

As a result of this exchange agreement, the expressions "foreign" and "foreign (non-U.S.)" are both used in this paper to denote "neither Canadian nor American".

REFERENCES

[1] Burchell, J.M.: "International Air Passenger Origin/Destination Project"; presentation to the Transportation and Communications Division of Statistics Canada, March 1981.

[2] "International Air Passenger O & D Statistics System - User Requirements Specifications"; prepared in the Aviation Statistics Centre (ASC) of the Transportation and Communications Division of Statistics Canada, revised March 1981.

[3] "International Air Passenger O & D Statistics System - Feasibility Study Report"; prepared in the ASC, January 1981.

[4] "International Air Passenger O & D Statistics System - Requirements for Functional Specifications (Concepts)"; prepared in the ASC, revised September 1981.

[5] Rosen, F.G. and Conroy, P.F.: "A Test of the Assignment Technique Based on a Survey of International Air Travellers at Montreal and Toronto"; a confidential report prepared for the Passenger and Aviation Economics Directorate of the Research Branch of the Canadian Transport Commission, February 1977.

[6] Carpenter, R.: "Specifications for Estimates of the Reliability for the Air Scheduled International Passenger Origin and Destination Estimates"; prepared in the Business Survey Methods Division of Statistics Canada, revised January 1982.

W-043F
c.4

MÉTHODOLOGIE DU SYSTÈME CANADIEN D'ESTIMATION DE L'ORIGINE ET DE LA DESTINATION DES PASSAGERS DES VOLS INTERNATIONAUX À HORAIRE FIXE

Greg Hunter et Lisa DiPiétro

Statistique Canada

INTRODUCTION

Selon le système d'estimation de l'origine et de la destination des passagers des vols internationaux à horaire fixe, communément appelé l'ASIPOD (Air Scheduled International Passenger Origin and Destination), on puise les données de deux enquêtes sur le trafic aérien pour produire des estimations du nombre de passagers sur les vols internationaux à horaire fixe, entre le Canada et les marchés étrangers, en fonction de diverses paires d'origines et de destinations. La première source, l'enquête sur l'origine et la destination des passagers payants, offre un échantillon de données sur les voyages internationaux où au moins un parcours d'un itinéraire est assuré par un transporteur canadien. Cette enquête, toutefois, présente des lacunes au niveau du champ d'observation, car elle ne fournit pas de données sur les voyages internationaux qui comportent les services d'un transporteur étranger pour l'ensemble de l'itinéraire. La deuxième source, l'enquête sur l'activité aéroportuaire, dénombre tous les passagers qui sortent du Canada ou qui y entrent d'un autre pays à bord d'un transporteur canadien ou étranger qui offre des services à horaire fixe, sans tenir compte de l'origine ou de la destination des passagers.

Ce document présente d'abord un aperçu des besoins des utilisateurs en matière de données sur l'origine et la destination des passagers sur les vols internationaux. On examine ensuite les principaux aspects des deux sources de données sur le trafic aérien, ainsi que le problème de la non-exhaustivité du dénombrement[1]. Enfin, ce texte explique de quelle façon le système d'estimation de l'ASIPOD permet de calculer des estimations du nombre de passagers, et les coefficients de variation correspondants, pour diverses paires d'origines et de destinations, que ces villes soient situées ou non dans les parties du marché international visées par l'enquête sur les passagers payants.

HISTORIQUE

Le remaniement du système d'estimation de l'ASIPOD est un des divers projets de révision du programme fédéral de la statistique de l'aviation. En 1979, Statistique Canada a entamé ce travail de révision en invitant le ministère des Transports et la Commission canadienne des transports (les deux "organismes utilisateurs") à former une équipe interministérielle, le Comité d'étude des besoins en statistiques sur l'aviation (CESA). Ce comité agit à titre de principal groupe d'étude pour l'équipe du projet ASIPOD; il se compose d'un chargé de projet, de représentants des ministères utilisateurs et

---

[1] This paper is also available in English[2].

de spécialistes en statistique mathématique, en analyse économique de l'aviation et en traitement des données. Jusqu'à présent, l'équipe du projet a défini les besoins des utilisateurs, effectué des études de faisabilité, établi les particularités des méthodes à appliquer et conçu le système de calcul.

BESOINS DES UTILISATEURS[1], [2]

L'équipe du projet ASIPOD a défini les besoins suivants des utilisateurs:

1) **Estimations par origine et destination des passagers des vols internationaux.**

Avant qu'un service commercial international à horaire fixe[3] à destination ou en provenance du Canada puisse être assuré, il doit exister un accord officiel entre le Gouvernement du Canada et le gouvernement du deuxième pays. Cet accord peut prendre la forme d'un échange provisoire de notes diplomatiques ou d'une entente globale sur le transport aérien conclue par voie de négociations.

Des négociations bilatérales sur les transports aériens nécessitent la participation de fonctionnaires des ministères canadiens des Affaires extérieures, des Transports et de l'Industrie et du Commerce, ainsi que de la Commission canadienne des transports. De plus, les pourparlers peuvent durer quelques mois ou quelques années.

Les routes aériennes pour les services à horaire fixe sont normalement l'objet principal des négociations, mais non le seul. Les articles contenus dans une entente sur le transport aérien peuvent concerner:

- le droit de survoler un certain territoire, ou d'y atterrir pour des raisons non liées au service normal;
- la désignation du transporteur qui peut desservir chaque route;
- le respect des lois et des règlements de chaque pays en matière d'entrée, d'autorisation, d'immigration, de passeports, de douanes et de quarantaine;
- la navigabilité, les certificats de compétence et les permis;
- l'échange de statistiques;
- les tarifs;
- les transferts de fonds;
- l'exemption d'impôts sur le revenu.

Pour discuter de ces sujets et surtout pour échanger des routes, les négociateurs doivent savoir où se trouvent les marchés du transport aérien et s'ils sont en expansion. En conséquence, ils doivent disposer des résultats d'analyses de coûts et de bénéfices, tant pour le Canada que pour les autres pays, qui découlent de l'attribution de diverses routes internationales à des transporteurs canadiens et étrangers. Afin de permettre au Canada de négocier

une juste part du marché, il est essentiel de posséder des estimations de l'origine et de la destination des passagers à bord des vols internationaux. A titre indicatif de l'importance de ces données pour l'économie canadienne, mentionnons que les recettes de l'ensemble des services aériens internationaux à destination ou en provenance du Canada ont atteint près de 2,3 milliards de dollars canadiens en 1980.

2) Objectifs de la révision du système d'estimation

Les quatre principaux objectifs visés par le projet de révision, dont le quatrième sera analysé en détail dans ce document, sont les suivants:

(i) Eliminer les problèmes signalés dans le système actuel.

Comme le système actuel ne prévoit pas d'imputation dans le cas des codes de transporteur inscrits de façon illisible sur le volet des billets d'avion, la catégorie "transporteurs inconnus" devient la troisième en importance dans les totalisations. En outre, il n'y a aucune vérification des codes qui permette de déterminer si un transporteur vole en direction ou à partir d'aéroports où il est effectivement autorisé à atterrir. Il faut donc prévoir la vérification et l'imputation des données illisibles ou fausses inscrites sur les volets, en plus des corrections qui seraient nécessaires à l'établissement des totalisations des vols exclusivement intérieurs.

D'autres erreurs d'observation ont été relevées, mais il est difficile de les corriger à l'aide d'un système d'estimation. Ce type d'erreurs inclut les cas où les transporteurs interprètent mal les instructions relatives aux volets; des erreurs fréquentes dans les numéros d'ordre des billets, qui sont utilisés pour le choix de l'échantillon; des erreurs dans les systèmes de traitement des transporteurs, etc. La rectification de ces erreurs d'observation, erreurs peu faciles à corriger, n'est pas un objectif du programme de révision.

(ii) Mettre au point un système de calcul simple, facile à utiliser et qui fournit des messages indicateurs sommaires.

Un grand volume d'enregistrements[4] doit être traité avec un minimum d'intervention manuelle. Etant donné qu'une multitude de passagers est répartie parmi un nombre considérable de paires d'origines et de destinations, les messages imprimés à chaque étape du système doivent résumer le traitement jusqu'à ce point, tout en indiquant les problèmes éventuels.

(iii) Fournir des tableaux des estimations du nombre de passagers sur les vols internationaux, soit régulièrement ou au besoin, de manière à simplifier les analyses que les utilisateurs doivent effectuer.

(iv) Produire des estimations, dont la fiabilité et la précision sont mesurables, sur l'origine et la destination des passagers à bord des vols internationaux à horaire fixe.

On a cru dans le passé que la précision de ces statistiques était très variable mais, en fait, leur fiabilité (c'est-à-dire la variabilité d'échantillonnage) était inconnue jusqu'à présent. L'estimation de la fiabilité de ces données permettra de mieux connaître la force des conclusions qu'on peut tirer de leur analyse. Les déductions faites sans une mesure de la fiabilité des données peuvent être en fait fort trompeuses.

FAISABILITÉ[3]

Les études de faisabilité effectuées dans le cadre de ce projet reposent sur le principe qu'il faut conserver la méthodologie fondamentale appliquée dans le système actuel. Par conséquent, la plupart de ces travaux ont visé à chercher des améliorations souhaitables, à les classer selon leur degré d'utilité et à déterminer lesquelles pouvaient être adoptées compte tenu des délais et du budget. Le maintien de la méthodologie fondamentale a fourni les grandes lignes du calcul des estimations du nombre de passagers des vols internationaux, mais non celui des estimations de la fiabilité de ces chiffres.

La population cible de ce système est l'ensemble des billets émis pour un voyage international (c'est-à-dire entre un pays étranger et le Canada ou les Etats-Unis). Ce type de billets est également l'univers de l'enquête sur l'origine et la destination des passagers payants, mais celle-ci vise uniquement les principaux transporteurs canadiens[5]. Chaque transporteur participant prend un coupon des billets dont le numéro d'ordre se termine par "0", tenant compte seulement des volets dont il est le premier à prélever. Ainsi, l'enquête porte sur un échantillon de 10% des volets pour des itinéraires qui incluent les services d'au moins un des transporteurs canadiens ou américains participants.

Quelques renseignements sur les marchés des transporteurs étrangers (non américains) peuvent être tirés de l'enquête sur l'origine et la destination des passagers payants. Par exemple, si un passager voyage par Air Canada d'Ottawa à Montréal, et de là à Paris par Air France, l'enquête sur l'origine et la destination des passagers payants comptera ce trajet international parce qu'un transporteur canadien intervient quelque part dans l'itinéraire. Le transporteur canadien doit déclarer tous les détails sur les transporteurs et les trajets, y compris la partie concernant Air France.

L'enquête sur l'origine et la destination des passagers payants ne tient pas compte cependant des volets où des transporteurs étrangers sont inscrits du début à la fin d'un itinéraire. Par exemple, si un passager suit un itinéraire sans correspondance qui le conduit de Paris à Montréal par Air France et de nouveau à Paris avec ce même transporteur, ce voyage ne figurera pas dans l'enquête sur l'origine et la destination des passagers payants. Toutefois ce genre d'itinéraire fait effectivement partie de la population cible de l'étude des voyages internationaux.

Ce dénombrement incomplet de la population cible constitue en fait un problème de non-exhaustivité pour le système d'estimation de l'ASIPOD.

Dans le système actuel, on utilise les données des enquêtes sur l'origine et la destination des passagers payants et sur l'activité aéroportuaire, et on applique une méthode appelée la technique d'attribution, afin de produire des estimations globales pour chaque marché.

L'enquête sur l'activité aéroportuaire procède à un dénombrement complet des passagers à bord des vols internationaux à destination ou en provenance de chaque aéroport canadien. Tous les transporteurs canadiens, américains et étrangers qui effectuent des vols à horaire fixe sont inclus, mais sans considération du premier point de départ ou de la destination finale des passagers.

Ainsi, l'enquête sur l'activité aéroportuaire donne un dénombrement du volume total des passagers de tous les transporteurs qui desservent la population cible. La technique d'attribution consiste à estimer le volume de passagers non dénombrés par l'enquête sur l'origine et la destination des passagers payants et à les classer dans des paires d'origines et de destinations. Avant d'expliquer plus en détail cette technique, il convient d'abord de décrire ces deux enquêtes sur le trafic aérien de façon un peu plus complète.

PRINCIPAUX ASPECTS DES DEUX ENQUÊTES

L'organisation de ces deux programmes d'enquête est confiée au Comité des transports aériens de la Commission canadienne des transports, en collaboration avec le ministère des Transports. Les données sont recueillies auprès des transporteurs aériens, pour le compte du Comité des transports aériens, par le Centre des statistiques de l'aviation (CSA) de Statistique Canada. En vertu des Règlements sur les transports aériens de la Loi sur l'aéronautique, les transporteurs sont obligés de remplir les états (questionnaires) du CSA.

Les statistiques sur l'origine et la destination des passagers payants parviennent au CSA grâce à l'état 35. Parmi les renseignements recueillis, on trouve entre autres:

- l'origine et la destination inscrites sur le billet;
- les points où un passager s'arrête pour prendre une correspondance sur la même ligne ou sur une ligne différente;
- le transporteur inscrit sur chaque volet.

Les renseignements sur l'origine et la destination des passagers payants sont fournis chaque mois par les principaux transporteurs canadiens qui émettent des billets à tarif unitaire et exploitent des services à horaire fixe. Depuis le 1er janvier 1982, les sept transporteurs canadiens qui participent à cette enquête sont Air Canada, CP Air, Eastern Provincial Airways, Nordair, Pacific Western Airlines, Air Ontario et Québecair. Les données américaines sont recueillies par le Civil Aeronautics Board auprès de tous les transporteurs autorisés, à l'exception de ceux qui assurent les transports par hélicoptère et les transports intérieurs en Alaska. Les chiffres pour chaque trimestre sont regroupés et les renseignements en double supprimés de façon à constituer un fichier des itinéraires complets, d'après l'origine et la destination du voyage complet ("TOD", Ticket Origin and Destination).

Par contre, les statistiques sur l'origine et la destination des passagers aériens sont traitées selon le concept de l'origine et de la destination du voyage simple ("DOD", Directional Origin and Destination), qui peut être défini comme "une suite de points de départ et de destination énumérés dans un

ordre qui indique la direction suivie dans l'itinéraire". Les voyages simples (DOD) sont les parties d'un itinéraire, lequel est divisé de façon que chaque voyage simple corresponde à une seule direction. Cette méthode exige qu'on décompose les itinéraires semi-circulaires ("open-jaw"), et les itinéraires de retour comme les itinéraires symétriques et circulaires, en voyages simples à sens unique. Pour obtenir les voyages DOD, un programme qui calcule des points de division est appliqué aux données sur l'origine et la destination des voyages complets. Ce procédé est automatisé dans le système qui regroupe les statistiques sur l'origine et la destination des passagers aériens, et il comporte le calcul de la distance entre divers points d'un itinéraire, suivi de comparaisons entre ces distances et la distance totale de l'itinéraire. En règle générale, on divise les itinéraires à partir du point le plus éloigné du lieu d'origine. Chaque voyage DOD ainsi formé est présenté de nouveau au programme jusqu'à ce qu'aucun nouveau point de division ne puisse être trouvé.

Les données sur l'activité aéroportuaire proviennent de l'état 32. Les caractéristiques relevées pour chaque vol incluent:

- le transporteur participant;
- l'aéroport déclarant;
- le point d'origine et la destination finale prévue pour le vol;
- le nom du dernier aéroport visité, dans le cas des arrivées, ou le prochain aéroport prévu à l'itinéraire, dans le cas des départs;
- le nombre de débarquements et d'embarquements de passagers payants.

Les données sur l'activité aéroportuaire sont envoyées tous les mois par les transporteurs transcontinentaux (Air Canada et CP Air) et régionaux (Eastern Provincial, Québecair, Nordair et Pacific Western Airlines) du Canada, par Norcanair et par tous les transporteurs étrangers (y compris les transporteurs américains) qui exploitent des services internationaux à horaire fixe à destination ou en provenance du Canada. Depuis le 1er janvier 1992, 10 transporteurs américains et 21 transporteurs étrangers remplissent une déclaration pour chaque aéroport canadien qu'ils desservent. Tout nouveau transporteur étranger qui obtient un permis d'exploitation de services à horaire fixe au Canada est automatiquement inclus dans le système de collecte des données sur l'activité aéroportuaire.

Les données sur le flux de passagers dans l'ensemble du trafic aérien sont extraites des renseignements sur l'activité aéroportuaire. Par "flux de passagers", on entend le nombre de personnes qui, pendant une période déterminée, voyagent à bord d'un transporteur donné entre un aéroport canadien déclarant et un point adjacent, c'est-à-dire le dernier aéroport visité ou le prochain point prévu à l'itinéraire. La technique d'attribution ne tient compte que des déclarations fournies par les transporteurs étrangers (non américains). Les éléments d'information extraits de cette enquête, qui sont utilisés pour déterminer les marchés formés par des paires d'origines et de destinations, sont le nombre de passagers payants embarqués ou débarqués au Canada, le transporteur au point d'entrée ou de sortie canadien et le nom du point d'entrée ou de sortie canadien. Par "point d'entrée ou de sortie canadien", nous entendons un aéroport qui participe à l'enquête et où un transporteur étranger (non américain) arrive au Canada ou en sort.

Cependant, d'après l'enquête sur l'origine et la destination des passagers payants, le point d'entrée/de sortie canadien pour les transporteurs canadiens et américains est le premier/dernier point canadien dans un itinéraire pour les vols à destination/en provenance du Canada. Quant aux transporteurs étrangers, le point d'entrée ou de sortie correspond au point à l'intérieur du Canada où un passager emprunte les services d'un transporteur étranger ou s'en sépare. Prenons l'exemple suivant:

Supposons qu'Air France assure un service Toronto - Montréal - Paris, et qu'il y a embarquement de certains passagers à Toronto et d'autres à Montréal.

Supposons également que deux suites de voyages simples sont possibles:

Winnipeg - Air Canada - Toronto - Air France - Montréal - Air France - Paris

Toronto - Air France - Montréal - Air France - Paris - British Airways - Londres

Le point de sortie canadien dans cet exemple serait Toronto, puisque c'est là que les passagers empruntent le transporteur étranger.

RÉSOLUTION DU PROBLÈME DE NON-EXHAUSTIVITÉ

Le problème de la non-exhaustivité du dénombrement sera examiné sous quatre aspects:

1) **Importance du problème**

Comme dans d'autres enquêtes, la non-exhaustivité du dénombrement pose un problème pour le système d'estimation de l'ASIPOD à cause de l'absence de données sur l'origine et la destination de la partie non dénombrée de la population cible.

Le tableau suivant donne un indice du volume de passagers qui se déplacent entre le Canada et neuf régions du monde. (Ces chiffres sont des estimations annuelles pour 1979. Les neuf régions du monde ne sont pas nommees parce que ces données sont confidentielles.)

Tableau 1 - Passagers des vols internationaux - estimations, 1979

Entre le Canada et...	Origine et destination des passagers payants	Non-dénombrement (le pourcentage du total est entre parenthèses)		Total
la région nº 1	116,050	495	(0.4)	116,545
la région nº 2	325,200	2,020	(0.6)	327,220
la région nº 3	99,010	12,628	(11.3)	111,638
la région nº 4	67,200	14,558	(17.8)	81,758
la région nº 5	575,010	126,076	(18.2)	703,086
la région nº 6	205,180	101,464	(33.1)	306,644
la région nº 7	1,221,040	908,876	(42.7)	2,129,916
la région nº 8	54,410	56,807	(51.1)	111,217
la région nº 9	45,330	57,267	(55.8)	102,597
Total mondial	2,708,430	1,282,191	(32.1)	3,990,621

D'après les pourcentages de non-dénombrement, il est évident que ce problème est important.

Si ce tableau était fait pour l'Est du Canada et les neuf mêmes régions du monde, on obtiendrait de plus hauts pourcentages de non-dénombrement pour chaque région du monde. Donc, il semble que, de façon générale, plus les régions géographiques sont désagrégées, plus le pourcentage de non-dénombrement est élevé. Pour citer un exemple, le pourcentage de non-dénombrement des passagers qui ont voyagé entre l'Est du Canada et la région nº 7 est de 55% contre 42,7% pour l'ensemble du Canada (tableau 1). Afin de clarifier l'affirmation qu'un haut degré d'agrégation géographique dans le choix des paires d'origines et de destinations produit généralement un faible pourcentage de non-dénombrement, il faut souligner que le non-dénombrement existe seulement dans le cas des vols des transporteurs étrangers qui prennent fin à un point d'entrée au Canada. Le dénombrement du trafic de passagers est exhaustif lorsque la partie canadienne des paires d'origines et de destinations n'est ni un point d'entrée, ni un point de sortie. Ainsi, lorsque le degré d'agrégation géographique augmente, on inclut plus de vols avec correspondance, ce qui fait diminuer le pourcentage de non-dénombrement.

2) La technique d'attribution[4]

La technique d'attribution consiste d'abord à estimer le volume de passagers non dénombrés, pour ensuite répartir ce volume dans des paires d'origines et de destinations.

A partir de l'enquête sur l'activité aéroportuaire, on effectue b comptages repères des passagers entrants et sortants aux aéroports canadiens, par transporteur. La valeur de b, le nombre de catégories auxquelles on peut attribuer un volume de voyageurs, se définit de la façon suivante:

$$b = 2 \times \sum_{i=1}^{g} n_i$$

où le "2" s'explique par le fait qu'il y a deux comptages, un pour les passagers qui entrent au Canada et un pour ceux qui sortent du pays,

g est le nombre d'aéroports canadiens qui servent de point d'entrée ou de sortie, et

n_i est le nombre de transporteurs étrangers (non américains) détenant un permis d'atterrissage, qui desservent le i^e aéroport canadien de point d'entrée ou de sortie.

L'enquête sur l'origine et la destination des passagers payants permet, par recoupement des transporteurs et des aéroports canadiens, d'obtenir un classement semblable des passagers entrants et sortants par origine et destination du voyage simple. Ainsi, l'échantillon de cette enquête fournit également b comptages.

A la première étape de la technique d'attribution, le volume de passagers non dénombrés, A_i, inclus dans le groupe d'attribution i, peut être estimé au moyen de l'équation suivante:

$$A_i = C_i - (1/f) \times D_i \quad (i = 1, \ldots, b)$$

où C_i est le volume total des passagers dans le groupe i, selon l'enquête sur l'activité aéroportuaire;

f est la fraction d'échantillonnage de l'enquête sur l'origine et la destination des passagers payants (c'est-à-dire 1/10);

D_i est le nombre échantillonné de passagers des vols internationaux dans le groupe i;

A_i correspond ainsi à l'estimation du nombre de passagers déplacés par un transporteur étranger et qui tombent dans le i^e des b groupes d'attribution, mais pour lesquels il manque des renseignements sur l'origine et la destination.

A l'étape suivante, on attribue le volume des passagers non dénombrés à des paires d'origines et de destinations. Ces paires reliées à la partie non dénombrée de la population cible, sont appelées des voyages DOD sans correspondance, puisqu'il s'agit de paires d'origines et de destinations de voyages simples assurés par des transporteurs étrangers qui n'ont pas de correspondances avec un transporteur participant canadien. Selon la technique d'attribution, on impute les voyages DOD sans correspondance au i^e groupe de la façon suivante:

(i) Une liste de tous les voyages DOD inclus dans D_i est dressée (c'est-à-dire au même aéroport, par le même transporteur étranger et dans la même direction).

(ii) Les trajets intérieurs sont éliminés (c'est-à-dire les trajets entre un point canadien et un point d'entrée ou de sortie canadien). (Ces trajets étant faits par un transporteur canadien, ils sont pris en compte dans l'enquête sur l'origine et la destination des passagers payants. Les voyages ainsi "tronqués" sont alors sans correspondance.)

(iii) Le volume des passagers non dénombrés, A_i, est réparti parmi les voyages DOD ainsi "échantillonnés", selon la proportion de ces voyages dans D_i.

De nouveaux totaux de passagers classés par origine et destination de voyages simples sont produits, et il s'agit alors de "passagers répartis".

La technique d'attribution repose sur l'hypothèse que les voyages DOD tronqués représentent bien le trafic sans correspondance, ce qui donne plus de poids à certaines paires d'origines et de destinations que dans l'enquête qui porte seulement sur l'origine et la destination des passagers payants. On peut donc trouver l'estimateur d_j^T, qui correspond au nombre total de passagers observés sur le marché pour la j^e paire d'origines et de destinations, en corrigeant la fraction d'échantillonnage de la manière suivante:

$$d_j^T = (1/f_j) \times d_j$$

$$\text{où } f_j = \frac{d_j}{(d_j/f) + a_j} \qquad (1)$$

et où f_j est la fraction d'échantillonnage corrigée pour la j^e paire d'origines et de destinations dans l'enquête sur l'origine et la destination des passagers payants;

d_j est le nombre de passagers des vols internationaux pour la j^e paire d'origines et de destinations dans l'échantillon de l'enquête sur l'origine et la destination des passagers payants, et

a_j est le nombre de passagers attribués à la j^e paire d'origines et de destinations.

Il convient de souligner que, conformément au point (iii) ci-dessus, a_j peut s'exprimer:

$$a_j = \frac{d_j}{D_i} \times A_i \qquad j \in i$$

où $\sum_{j \in i} d_j = D_i$

et $\sum_{j \in i} a_j = A_i$

3) Un exemple d'application de la technique d'attribution

Prenons un exemple simple pour illustrer comment fonctionne la technique d'attribution. Les données sur l'activité aéroportuaire fournies par British Airways révèlent que 120 passagers sont embarqués à Montréal à destination de Londres. Dans ce cas, C_i = 120. Supposons que l'enquête sur l'origine et la destination des voyageurs payants n'indique que deux suites de voyages DOD:

Nombre de passagers dans l'échantillon	Estimation du nombre de passagers	DOD
1	10	YWG-AC-YMX-BA-LON-LO-WAW
2	20	YYZ-CP-YMX-BA-LON-LH-HAM

où les codes doivent être interprétés comme suit:

Code	Désignation
YWG	Winnipeg
YMX	Montréal (Mirabel)
LON	Londres
WAW	Varsovie
YYZ	Toronto
HAM	Hambourg
AC	Air Canada
BA	British Airways
LO	LOT
CP	CP Air
LH	Lufthansa

Ainsi, $D_i = 3$ et $A_i = 120 - (1/.1) \times 3 = 90$.

Les voyages DOD tronqués, la proportion de ces voyages inclus dans D_i, le nombre de passagers à répartir et les estimations totales pour le marché sont donc:

Voyages DOD tronqués	Proportions	Passagers à répartir (a_i)
YMX - BA - LON - LO - WAW	1/3	30
YMX - BA - LON - LH - HAM	2/3	60

4) Lacunes de la technique d'attribution

Cette technique comporte toutefois certaines lacunes.

La technique d'attribution est fondée sur le postulat que les voyages DOD tronqués représentent bien le trafic sans correspondance. Examinons l'exemple hypothétique que nous venons d'illustrer afin de voir si cette hypothèse est, par intuition, raisonnable. La technique d'attribution suppose que les passagers qui partent de Montréal à bord d'un vol d'un transporteur étranger particulier auront la même destination finale que les passagers dont l'origine est Winnipeg ou Toronto et qui passent par Montréal. Cela revient à utiliser le comportement des voyageurs des communautés ethniques, comme les groupes polonais de Toronto et allemands de Winnipeg, pour tracer le comportement des communautés à prédominance française de Montréal. En outre, le caractère douteux du postulat de similitude entre les voyages avec et sans correspondance a été démontré empiriquement dans une étude-pilote [5] effectuée par la Commission canadienne des transports, en 1977. Bref, il existe des arguments non seulement intuitifs, mais aussi empiriques qui réfutent le postulat de base de cette technique.

La précision des estimations du nombre de passagers par paire d'origines et de destinations est mise en cause par toute exception à l'hypothèse que les voyages DOD tronqués représentent bien le trafic sans correspondance. Or, comme dans le cas du tableau 1, de grands volumes de passagers sont attribués à des paires d'origines et de destinations à partir d'un échantillon "réel" de petite taille. Par exemple, la fraction d'échantillonnage "réelle" du marché formé entre le Canada et la région n° 7 n'est pas 10%, comme dans l'enquête sur le trafic intérieur, mais 5.7% (c'est-à-dire (100% - 42.7%) x 10%), à cause de l'exclusion du trafic sans correspondance de l'univers. Par conséquent, un échantillon de moins de 10% sert à répartir un grand volume de passagers. L'existence d'exceptions à l'hypothèse en question pourrait donc causer de grands biais dans les estimations.

Comme les chiffres tirés de l'enquête sur l'activité aéroportuaire constituent des repères sur le volume du trafic, leur précision est très importante. Bien qu'aucune évaluation n'ait été entreprise concernant l'ampleur du biais attribuable à des erreurs d'observation dans les chiffres sur l'activité aéroportuaire, les économistes qui analysent les statistiques sur l'aviation croient que de telles erreurs dans cette enquête sont petites. Toutefois, des consultations sont actuellement en cours avec les grands transporteurs aériens canadiens pour trouver des façons de réduire les exigences des organismes gouvernementaux en matière de déclarations. Au terme de ces discussions, l'activité aéroportuaire pourrait devenir l'objet d'une

enquête-échantillon. Lorsque viendra le moment d'établir le plan de sondage, il faudra tenir compte de l'importance d'obtenir des données précises sur le volume des passagers aux points d'entrée ou de sortie.

La technique d'attribution suppose que tout le trafic des passagers non dénombrés est pris en compte dans le trafic sans correspondance. Une façon intéressante de vérifier cette hypothèse serait de comparer, pour une même période de référence, le nombre de passagers d'un transporteur canadien déclaré dans l'enquête sur l'activité aéroportuaire et une estimation faite à partir de l'échantillon de l'enquête sur l'origine et la destination des voyageurs. Cette estimation correspondrait au nombre total de passagers dont les voyages DOD comportent le même transporteur pour le trajet international et le même aéroport canadien comme point d'entrée ou de sortie. Il faudrait pouvoir dire si les écarts entre le chiffre déclaré et les estimations proviennent de différences entre les notions appliquées par ces enquêtes et, dans l'affirmative, vérifier si ces différences sont prises en considération dans le système d'estimation de l'ASIPOD. Si le système n'en tient pas compte, cela pourrait confirmer que l'hypothèse selon laquelle tout le trafic des passagers non dénombrés est compris dans le trafic sans correspondance soulève des problèmes.

L'étude d'un bon nombre de ces lacunes devrait donc être poussée. Le système actuel comporte de sérieuses lacunes, mais on n'a pas encore réussi à trouver une solution de rechange à la technique d'attribution, étant donné les contraintes de temps et de coût. Néanmoins, les améliorations apportées aux estimations du trafic pour chaque marché par le système d'estimation de l'ASIPOD seront appréciables.

L'application de la technique d'attribution à l'estimation des marchés internationaux représente une solution nouvelle pour un problème de taille. Elle ne règle pas complètement les difficultés causées par la non-exhaustivité du dénombrement, mais elle constitue un grand pas en avant, comme d'ailleurs la production de statistiques sur la variance des estimations. Les chiffres sur la variance devraient à la fois tenir compte de la technique d'attribution et de ses hypothèses, et fournir une mesure claire de la fiabilité des estimations relatives aux DOD.

ESTIMATION DE LA VARIANCE[6]

L'estimateur de la variance des estimations du trafic international par paires d'origines et de destinations est une modification simple de l'estimateur de la variance utilisé dans l'enquête sur l'origine et la destination des passagers payants.

1) Variance de l'estimation du trafic avec correspondance

Le calcul de l'estimateur de la variance des estimations de l'origine et de la destination des passagers payants dépend du nombre de passagers que chaque billet attribue aux paires d'origines et de destinations (c'est-à-dire au domaines d'intérêt). Rappelons que chaque billet est choisi avec une probabilité de 0,1 et que l'itinéraire peut être divisé en plusieurs segments ou voyages DOD. Un billet peut faire inscrire 0, 1, 2, etc. passagers dans un domaine d'intérêt donné. Par exemple, l'itinéraire

YWG - AC - LON - BA - YYZ

serait coupé, selon le programme qui calcule les points de division, en deux voyages simples:

YWG - AC - LON
et
LON - BA - YYZ.

Prenons les estimations du nombre total de passagers sortants et entrants, sans distinguer la direction du voyage. Parmi les diverses paires d'origines et de destinations, le billet en question ajouterait des passagers aux domaines suivants:

Domaine d'intérêt	Nombre de passagers
Winnipeg - Londres	1
Toronto - Londres	1
Canada - Londres	2
Canada - Europe	2
Est du Canada - Europe	1
Ouest du Canada - Europe	1

Notons que le nombre de passagers par billet dépend du niveau de répartition géographique et, par conséquent, de l'origine et de la destination (c'est-à dire des domaines d'intérêt) pour lesquelles des estimations du trafic sont calculées.

On peut trouver $\hat{d}$, l'estimation du nombre de passagers qui voyagent entre une origine et une destination en particulier, d'après l'enquête sur l'origine et la destination des passagers payants, à l'aide de la formule suivante:

$$\hat{d} = \sum_{i=1}^{n} \frac{x_i}{f}$$

où x_i est le nombre de voyages simples appartenant à une paire d'origines et de destinations pour le i^e billet.

n est le nombre total de billets dans l'échantillon, et

N est le nombre total de billets dans la population de l'enquête sur l'origine et la destination des passagers payants.

f est la fraction d'échantillonnage, c'est-à-dire, n/N = 0,1.

L'enquête sur l'origine et la destination des passagers payants utilise effectivement un échantillon aléatoire simple de 10%, puisque

(i) le choix des volets dont le numéro d'ordre se termine par "0" produit un échantillonnage systématique, et

(ii) la distribution des billets ne comporte aucun cycle qui causerait une relation entre le nombre de passagers estimé par l'enquête et le dernier chiffre du numéro d'ordre.

L'estimation de la variance peut donc s'écrire:

$$\hat{var}(\hat{d}) = N^2 \frac{1}{n} (1-f) v_s$$

où $v_s = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

et $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$,

et le coefficient de variation peut s'exprimer sous la forme:

$$\hat{cv}(\hat{d}) = \sqrt{\hat{var}(\hat{d})}/\hat{d} \quad .$$

A noter que var(d) peut s'écrire également:

$$var(d) = \frac{(1-f)}{f^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (2)$$

où n est supposé suffisamment grand pour que n/(n-1) égale approximativement 1.

2) Variance des estimations globales relatives à un marché

La méthode de calcul des coefficients de variation pour l'ensemble d'un marché doit assurer que les données imputées sont une fonction des données de l'échantillon. Autrement dit, des échantillons différents produisent des valeurs différentes des données imputées et, en conséquence, des grandeurs différentes des estimations globales relatives à un marché. La réutilisation de certaines parties de l'échantillon a un effet sur la distribution d'échantillonnage des estimations.

La méthode que nous proposons corrige la fraction d'échantillonnage de 10% pour qu'elle corresponde au pourcentage de passagers réellement échantillonnés dans chaque domaine d'intérêt. Comme on prend pour acquis l'application de la technique d'attribution, on suppose également que les voyages DOD tronqués représentent bien le trafic sans correspondance. Notre mesure de fiabilité constitue une mesure de précision seulement si cette hypothèse est valable. Il est donc raisonnable de rectifier les poids des voyages DOD échantillonnés afin de prendre en considération le trafic sans correspondance. Ainsi, la fraction d'échantillonnage nécessaire pour l'estimation du j^e domaine d'intérêt serait f_j, comme définie dans l'équation (1) plus haut. Dans l'équation (2), f_j remplacerait f pour le calcul de var(d), ce qui donne la formule pour la variance des estimations globales relatives à un marché.

$$\hat{var}(\hat{d}_j) = \frac{(1-f_j)}{f^2_j} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Ensuite, le coefficient de variance des estimations globales relatives à un marché serait:

$$cv(\hat{d}_j^T) = \sqrt{var(\hat{d}_j^T)} \; / \; \hat{d}_j^T$$

Signalons que $f_j < 0{,}10$ lorsque $a_j > 0$, ce qui signifie que la fraction d'échantillonnage est inférieure à 10% quand il est nécessaire d'inclure des données imputées dans l'estimation du nombre global de passagers dans un marché. Ainsi, l'inclusion de f_j au lieu de 0,10 dans l'expression de la variance d'une estimation produit un coefficient ajusté qui s'applique aux estimations globales relatives à un marché.

Cette méthode tient compte du fait que des données d'échantillonnage sont utilisées dans la technique d'attribution, mais elle est fondée, comme le procédé qui détermine les estimations elles-mêmes, sur l'hypothèse que les voyages DOD tronqués représentent bien le trafic sans correspondance.

PROJETS À L'ÉTUDE

Le besoin de disposer de renseignements sur les marchés des transports aériens pour les négociations bilatérales sur les services internationaux a été expliqué dans une des sections précédentes. L'échange de statistiques constitue un des éléments négociables prévus dans les accords sur le transport aérien, et des ententes permettant de tels échanges existent déjà entre le Canada et plusieurs pays. Les notions que ces pays appliquent dans leurs enquêtes ainsi que la qualité de leurs données nous laissent croire que ces chiffres pourraient être utilisés dans le système d'estimation de l'ASIPOD. De telles données fourniraient un échantillonnage des trajets non dénombrés dans l'univers des billets émis pour les vols internationaux. Des études de faisabilité actuellement en cours tentent de déterminer si le nombre de pays qui pourraient échanger ces statistiques contribuerait à améliorer la précision d'un nombre d'estimations suffisant pour justifier l'expansion du système d'estimation de l'ASIPOD en vue d'inclure des "données échangées".

RÉSUMÉ

Le système d'estimation de l'ASIPOD est fondé sur les données de deux enquêtes sur le trafic aérien et permet de produire des estimations du nombre de passagers à bord des vols internationaux en fonction de l'origine et de la destination. La technique d'attribution offre une solution au problème du non-dénombrement du trafic sans correspondance. Les hypothèses sur lesquelles reposent cette technique soulèvent assez de doutes pour qu'une évaluation du biais des estimations mérite d'être entreprise. Cependant, le nouveau système comportera des améliorations importantes qui diminuent le biais des estimations, tout en calculant un estimateur de la fiabilité des résultats. Par conséquent, les négociations bilatérales sur le transport aérien international nous permettront de mieux déterminer la valeur des conclusions tirées au sujet des marchés des transports aériens à partir de ces estimations.

REMERCIEMENTS

Les auteurs remercient leurs collègues de la Division des transports et des communications et de la Division des méthodes d'enquêtes-entreprises de Statistique Canada pour les nombreuses remarques pertinentes faites pendant la rédaction de ce document.

RENVOIS

1 Il s'agit d'un problème de "non-dénombrement" plutôt que de "sous-dénombrement" parce qu'il est impossible d'inclure une grande partie de l'univers dans le cadre de l'enquête sur le trafic international.

2 To obtain a copy of the English version, please write to:

Greg Hunter,
Business Suvey Methods Division,
Jean Talon Building, 5th Floor, Section C8,
Tunney's Pasture,
Ottawa, Ontario, Canada.
K1A 0T6

3 Un service aérien commercial international à horaire fixe est défini comme l'exploitation de vols entre des points situés au Canada et des points à l'extérieur pour le transport de passagers, de biens ou de courrier par aéronef selon un horaire fixe et à tarif unitaire. Ce genre de service s'appelle un service à tarif unitaire.

4 Il y a quelque 625,000 enregistrements d'origine et de destination par année, et quelque 820,000 enregistrements concernant l'activité aéroportuaire par année.

5 Il existe un programme d'échange de données sur l'origine et la destination des passagers entre le Canada et les États-Unis, dans lequel les États-Unis fournissent au Canada des renseignements sur les itinéraires complets où:

(i) un point situé au Canada et un point situé aux États-Unis figurent sur le billet, ou

(ii) un transporteur des États-Unis s'est rendu en un point situé au Canada ou a quitté un point canadien, ou

(iii) un transporteur canadien s'est rendu en un point situé aux États-Unis ou a quitté un point américain.

A cause de cet accord, nous employons les expressions "étranger" ou "étranger (non américain)" pour désigner ce qui est "ni canadien, ni américain".

BIBLIOGRAPHIE

[1] Burchell, J.M.: "International Air Passenger Origin/Destination Project"; exposé présenté à la Division des transports et des communications de Statistique Canada, mars 1981.

[2] "International Air Passenger O & D Statistics System - User Requirements Specifications"; document du Centre des statistiques de l'aviation (CSA) de la Division des transports et des communications de Statistique Canada, revu en mars 1981.

[3] "International Air Passenger O & D Statistics System - Feasibility Study Report"; document du CSA, janvier 1981.

[4] "International Air Passenger O & D Statistics System - Requirements for Functional Specifications (Concepts)"; document du CSA, revu en septembre 1981.

[5] Rosen, F.G. and Conroy, P.F.: "A Test of the Assignment Technique Based on a Survey of International Air Travellers at Montreal and Toronto"; rapport confidentiel établi pour la Direction générale des études économiques du transport de passagers et de l'aviation, Direction de la recherche, Commission canadienne des transports, février 1977.

[6] Carpenter, R.: "Specifications for Estimates of the Reliability for the Air Scheduled International Passenger Origin and Destination Estimates"; document de la Divison des méthodes d'enquêtes-entreprises de Statistique Canada, revu en janvier 1982.

POOR

QUALITY

ORIGINAL